Claims

We claim:

1   1. A method for detecting highlights from videos, comprising:

2   extracting audio features from the video;

3   classifying the audio features as labels;

4   extracting visual features from the video;

5   classifying the visual features as labels; and

6   fusing, probabilistically, the audio labels and visual labels to detect

7   highlights in the video.

1   2. The method of claim 1, in which the video is compressed.

1   3. The method of claim 1, in which silent features are classified according to audio

2   energy and zero cross rate.

1   4. The method of claim 1, in which the audio features are Mel-scale frequency

2   cepstrum coefficients.

1   5. The method of claim 1, in which the audio features are MPEG-7 descriptors.

1   6. The method of claim 1, in which the audio features are classified using Gaussian

2   mixture models.

11

1   7. The method of claim 1, in which the audio labels are selected from the group

2   consisting of applause, cheering, ball hit, music, male speech, female speech, and

3   speech with music.

1   8. The method of claim 1, in which the visual features are based on motion activity

2   descriptors.

1   9. The method of claim 1, in which the visual features include dominant color and

2   motion vectors.

1   10. The method of claim 1, in which a variance of the motion activity is quantized

2   to obtain the visual labels.

1   11. The method of claim 1, in which the motion activity is averaged to obtain the

2   visual labels.

1   12. The method of claim 1, in which the visual labels are selected from the group

2   consisting of close shot, replay, and zoom.

1   13. The method of claim 1, in which the probabilistic fusion uses a discrete-

2   observation coupled hidden Markov model.

1   14. The method of claim 13, in which the discrete-observation coupled hidden

2   Markov model includes audio hidden Markov models and visual hidden Markov

3   models.

1   15. The method of claim 14, in which the discrete-observation coupled hidden

2   Markov model is generated from a Cartesian product of states of the audio hidden

3   Markov models and the visual hidden Markov models, and a Cartesian product of

4   observations of the audio hidden Markov models and the visual hidden Markov

5   models.

1   16. The method of claim 13, further comprising:

2   training the discrete-observation coupled hidden Markov model with hand

3   labeled videos.

1   17. The method of claim 1, in which the video is a sport video.

1   18. The method of claim 1, further comprising:

2   determining likelihoods for the highlights; and

3   thresholding the highlights.

1   19. The method of claim 1, in which the audio portion of the video is compressed.

1   20. The method of claim 1, in which the visual portion of the video is compressed.